



**Українська Федерація Інформатики
Інститут кібернетики імені В. М. Глушкова НАН України
Вищий навчальний заклад Укоопспілки
«ПОЛТАВСЬКИЙ УНІВЕРСИТЕТ ЕКОНОМІКИ І ТОРГІВЛІ»
(ПУЕТ)**

ІНФОРМАТИКА ТА СИСТЕМНІ НАУКИ (ІСН-2015)

**МАТЕРІАЛИ
VI ВСЕУКРАЇНСЬКОЇ НАУКОВО-ПРАКТИЧНОЇ
КОНФЕРЕНЦІЇ ЗА МІЖНАРОДНОЮ УЧАСТЮ**

(м. Полтава, 19–21 березня 2015 року)

За редакцією професора О. О. Ємця

**Полтава
ПУЕТ
2015**

ОНТОЛОГИЧЕСКИЙ АНАЛИЗ ДАННЫХ МОНИТОРИНГА АВАРИЙНОГО СОСТОЯНИЯ ШАХТНОЙ АТМОСФЕРЫ

В. В. Слесарев, д. т. н., профессор,

В. С. Маргитич, студент

Государственный ВУЗ «Национальный Горный Университет»

victormargitich@gmail.com

Горное производство является сложной и опасной отраслью. В ходе ведения работ могут возникать различные опасные происшествия природного и технологического характера. Поэтому проблемы безопасности на объектах горного производства имеют особое значение. Анализ и оценка опасности возможных аварий является одной из ключевых проблем промышленной безопасности. Для комплексного охвата состояния горного производства применяют системы контроля и управления, например, такие, как УТАС. (г. Донецк). Выявление классов проблемных ситуаций осуществляется одним из методов интеллектуального анализа данных – кластеризацией. Цель кластерного анализа заключается в создании групп сходных (однородных) объектов. В то же время его действие состоит в привнесении структуры в анализируемые данные. Кластерный анализ является эвристической процедурой, не имеющей статистического обоснования, но даёт возможность произвести своеобразную разведку – изучить «структуру совокупности». Он применяется в случае, если нет априорных гипотез относительно классов на описательной стадии исследования. Разные кластерные методы могут порождать и порождают различные решения для одних и тех же данных.

В качестве дескрипторов (A'') для кластерного анализа характеристик аварийного процесса выбраны следующие параметры: окись углерода, углекислый газ, температура, метан. Для выявления непересекающихся подклассов и установления

значений признаков необходимо решить задачу классификации прецедентов – пожаров, имевших место на угольных шахтах. Предположим, что можно разбить множество значений признаков, характеризующих проблемную ситуацию в реализованных решениях (управляющих воздействиях), обозначенное A^n , на непересекающиеся подмножества, соответствующие классам проблемной ситуации $(a_{i1}, \dots, a_{im}, Category_j)$. То есть была подготовлена выборка данных из протоколов успешно ликвидированных аварий на шахтах в виде:

$$A^n = \{A_i, \quad i = 1, \dots, n\};$$

$$A_i = (a_{i1}, \dots, a_{ij}, \dots, a_{im});$$

$$\psi: A_i \rightarrow Category_k;$$

$$\forall A_i \in A^n \quad \exists category_k \in Category: \psi(A_i) = category_k,$$

где A^n – множество данных протоколов ликвидации аварий;

A_i – вектор значений данных;

a_i – аварийный параметр;

$Category_k = (category_1, category_2, \dots, category_j)$ – множество классов аварийных ситуаций.

Распознавание аварийных ситуаций состоит в том, чтобы на основании мониторинга объекта и анализа вектора признаков $a[i, j]$ (данные в определённый момент времени) отнести его к определённому классу возможных аварийных ситуаций в случае, если производится оценка реализуемых решений (категорий) $cat_j \in Cat$. И уже наступившей аварийной ситуации (в случае если анализируются характеристики реализуемых решений). Далее полученный класс соотносится с характерным для него управляющим воздействием $b[i, j]$ и выдаётся рекомендации руководителю аварийных работ.

Исследования проводились в соответствии с методологией кластерного анализа, включающей пять основных этапов:

- 1) отбор выборки для кластеризации;

- 2) определение множества признаков, по которым будут оцениваться объекты в выборке, и способы их стандартизации;
- 3) вычисление той или иной меры сходства между объектами;
- 4) применение метода кластерного анализа для создания групп сходных объектов;
- 5) проверка достоверности результатов кластерного анализа.

Для реализации кластерного анализа данных использовался модифицированный пакет SPSS 11.5. На начальном этапе проведения кластерного анализа неизвестно количество классов в онтологии прецедентов оперативных планов ликвидации аварий (ОПЛА). Для определения количества кластеров (то есть классов ОПЛА) применяется иерархический агломерационный алгоритм. В результате количество кластеров становится равным $(N - 1)$. Процедура повторяется, пока все кластеры не объединятся. На любом этапе объединение можно прервать, получив нужное число кластеров. Выбор метода образования кластеров, метода расчёта дистанционной меры и меры подобия определяется с экспертом и учётом того, что рассматриваемые признаки представляют собой интервальные значения. В работе это метод «связь между группами» (*Between groups linkage*). Дистанция между кластерами равняется среднему значению дистанций между всеми возможными парами наблюдений. Информация, необходимая для расчёта дистанции, находится на основании всех возможных пар наблюдений. В качестве дистанционной меры выбран квадрат «евклидового» расстояния (*squred Euclidean distance*):

$$dist = \sum_{i,j=1}^n (x_i - y_j)^2.$$

Значение признаков стандартизуются модифицированными средствами SPSS 11.5. Для определения того, какое количество кластеров следовало бы считать оптимальным, была проанализирована агломерационная матрица, построенная на основании дистанционной меры – квадрат евклидового расстояния и стандартизованных значений признаков. Для этого после определения кластеров был проведён анализ данных

методом k «средних». Он проводился на основании предшествующих исследований. Первые наблюдения, содержащиеся в файле с данными, используются как первый кластер. Вычисляется расстояние между центрами кластеров и каждым объектом, и объект приписывается к тому кластеру, к которому он ближе всего. Завершив приписывание, по алгоритму вычисляются средние значения для каждого кластера. Набор k «средних» представляет собой координаты нового положения кластера. По алгоритму вновь вычисляем расстояние от каждого объекта до центра кластеров и приписываем объекты к ближайшему кластеру. Вновь вычисляются центры тяжести кластеров, и этот процесс повторяется до тех пор, пока центры тяжести не перестанут «мигрировать» в пространстве. На рисунке изображено распределение данных мониторинга трёх параметров в исходящей струе аварийных зон: окиси углерода, углекислого газа, температуры. На этом рисунке наблюдается распределение трёх групп наблюдений, что подтверждает возможность классификации аварийных ситуаций на шахтах методом кластеризации. Окончательные значения кластеров были получены уже на третьей итерации. С помощью метода k – средних было подтверждено, что исходное множество данных мониторинга шахтной атмосферы в зоне пожара было верно разбито на три предопределённых класса, классы получились достаточно однородными и равномошными.

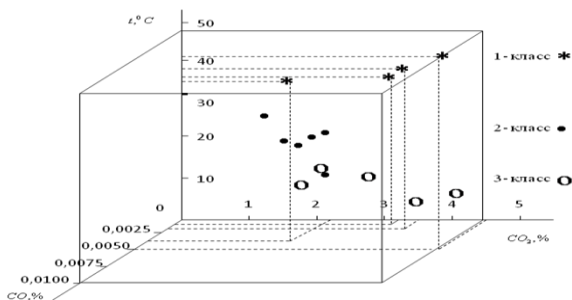


Рис. 1. Пространственное распределение данных мониторинга шахтной атмосферы в предаварийном состоянии

Проанализировав с экспертами результаты кластеризации: выявлены закономерности и были разработаны правила, позволяющие по значениям признаков отнести аварийную ситуацию (АС) к одному из трёх классов.

Научная значимость работы. Анализируя поступающую информацию о параметрах окиси углерода, углекислого газа, температуры и метана на аварийном участке, а также определяя текущую ситуацию внутри системы, выявлены скрытые количественные характеристики отношений между экземплярами и классами аварий. В соответствии с предложенной методикой формирования онтологий прецедентов, проведён кластерный анализ данных мониторинга процессов ликвидации пожаров на шахтах. На основании выделенных групп с экспертами были сформулированы правила распознавания классов проблемных ситуаций.